

随机信号处理笔记

宋佳欢

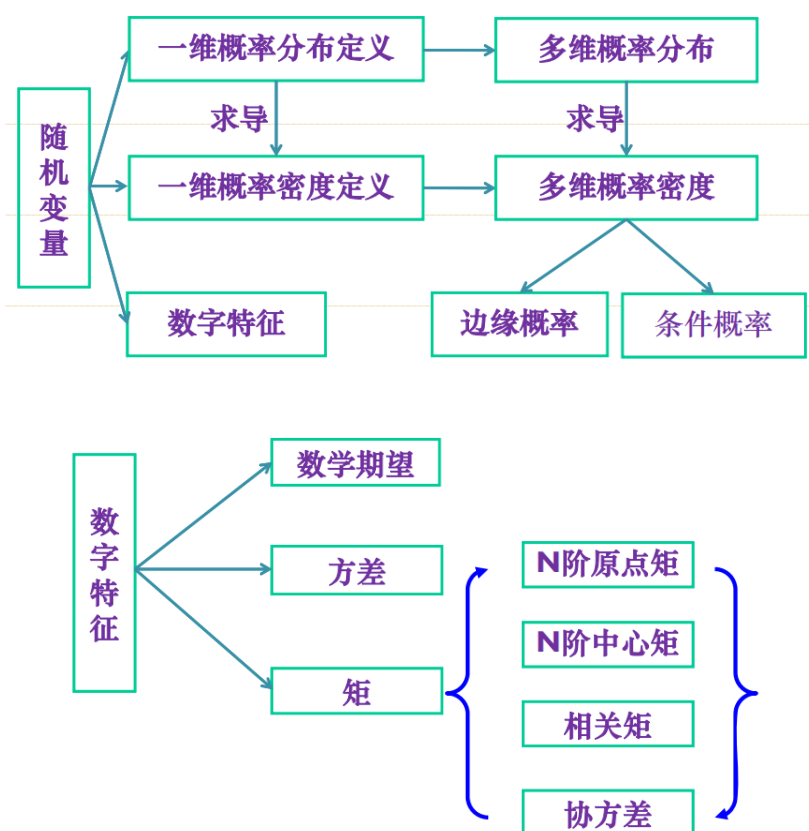
2019 年 10 月 11 日

目录

1 随机信号基础	2
1.1 随机变量与随机过程	2
1.1.1 随机过程的数学定义	2
1.1.2 随机过程的分类	2
1.2 随机信号的时域（统计）表示	3
1.2.1 随机信号的一维概率分布	4
1.2.2 随机信号的 N 维概率分布	4
1.2.3 离散时间随机过程的数字特征	4
2 维纳滤波器	6
2.1 正交性原理	6
2.2 最小均方误差	7
2.3 维纳-霍夫方程（求解）	8
3 最速下降法	10
3.1 最速下降法的基本思想	10
3.2 最速下降算法应用于维纳滤波器	10
3.3 最速下降法的稳定性	10
4 自适应滤波器	12
4.1 LMS 自适应滤波算法	13
4.2 LMS 算法的收敛性证明	13

1 随机信号基础

1.1 随机变量与随机过程



1.1.1 随机过程的数学定义

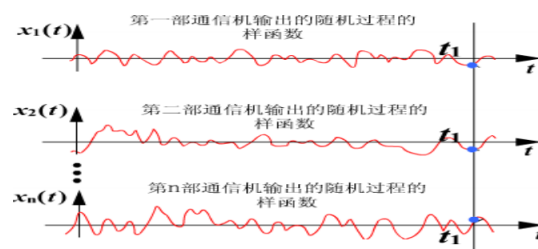
定义 1: 设 E 为随机试验, 它的样本空间是 $S = \{\xi\}$, 对于每一个 $\xi \in S$, 总有一个确定的时间函数 $u(t, \xi)$ 与之对应。这样可得到一族时间 t 的函数, 该簇称为随机过程。簇中的每一个函数称为这个随机过程的样本函数。

定义 2: 设 E 有一个过程 $u(t)$, 对于每一个时刻 $t_j (j = 1, 2, \dots)$, $u(t_j)$ 是一个随机变量, 则称 $u(t)$ 为随机过程。(默认采用该定义描述)

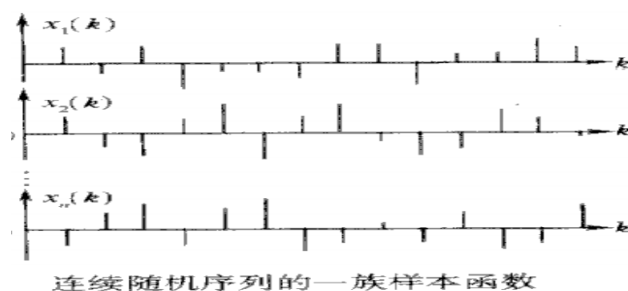
- A. 当 t 和 ξ 都是可变量时, u 是一个时间函数族;
- B. 当 t 是可变量、 ξ 固定时, u 是一个确定的时间函数;
- C. 当 t 固定, ξ 是可变量时, u 是一个随机变量;
- D. 当 t 固定, ξ 固定时, u 是一个确定数。

1.1.2 随机过程的分类

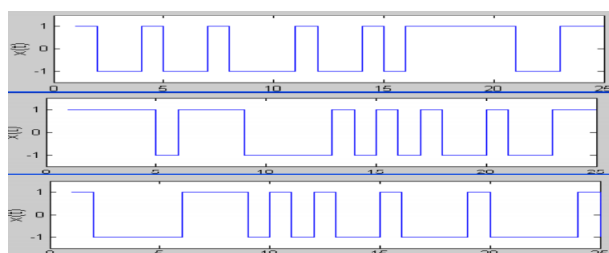
1. 连续型随机过程: 时间、状态都连续。



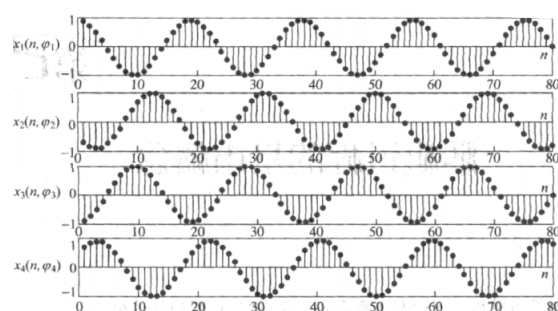
2. 离散时间型随机过程：时间离散，状态连续。



3. 离散幅度型随机过程：时间连续，状态离散。



4. 离散型随机过程：时间、状态都离散。



(数字信号处理，处理的是第 2 种信号)

1.2 随机信号的时域 (统计) 表示

考虑由时间序列 $u(n), u(n-1), \dots, u(n-M+1)$ 示的离散时间随机过程 $\{u(n)\}$ 。为了简化表示方式，从现在起，也使用 $u(n)$ 表示该离散时间随机过程。(随机过程等价于多维随机变量)

1.2.1 随机信号的一维概率分布

对于一个随机过程 $u(n)$ ，在任意时刻 n 是一个随机变量，它的一维概率分布函数为：

$$F_{u(n)}(u, n) = P\{u(n) \leq u\}$$

若 $F(u, n)$ 的一阶偏导数存在，则随机过程 $u(n)$ 的一维概率密度函数为：

$$f_{u(n)}(u, n) = \frac{\partial F(u, n)}{\partial u}$$

1.2.2 随机信号的 N 维概率分布

对于一个随机过程 u ，在任意 N 个时刻 n_1, n_2, \dots, n_N ，可构成 N 维随机变量 $\{u(n_1), u(n_2), \dots, u(n_N)\}$ ，它的的 N 维联合概率分布函数为：

$$F_u(u_1, u_2, \dots, u_N; n_1, n_2, \dots, n_N) = P\{u(n_1) \leq u_1, u(n_2) \leq u_2, \dots, u(n_N) \leq u_N\}$$

若 $F_u(u_1, u_2, \dots, u_N; n_1, n_2, \dots, n_N)$ 对 u_1, u_2, \dots, u_N 的偏导数存在，则随机过程 $u(n)$ 的 N 维概率密度函数为：

$$f_u(u_1, u_2, \dots, u_N; n_1, n_2, \dots, n_N) = \frac{\partial^2 F_u(u_1, u_2, \dots, u_N; n_1, n_2, \dots, n_N)}{\partial u_1 \partial u_2 \dots \partial u_N}$$

1.2.3 离散时间随机过程的数字特征

随机过程的数字特征有随机变量的数字特征推广而来，但一般不再是确定的数值，而是关于时间的函数。

1. 数学期望：表示随机过程所有样本函数的统计平均函数。

$$\mu(n) = E(u(n)) = \int_{-\infty}^{+\infty} u f_{u(n)}(u, n) du$$

2. 自相关函数

复数域： $r(n, n-k) = E[u(n)u^*(n-k)]$

实数域： $r(n, n-k) = E[u(n)u(n-k)]$

$$k = 0, \pm 1, \pm 2, \dots$$

3. 自协方差函数

复数域：

$$c(n, n-k) = E\{[u(n) - \mu(n)][u(n-k) - \mu(n-k)]^*\}$$

实数域：

$$c(n, n-k) = E\{[u(n) - \mu(n)][u(n-k) - \mu(n-k)]\}$$

$$k = 0, \pm 1, \pm 2, \dots$$

分别减去了均值。

4. 自相关函数与自协方差函数的关系

复数域：

$$c(n, n-k) = r(n, n-k) - \mu(n)\mu^*(n-k)$$

实数域：

$$c(n, n-k) = r(n, n-k) - \mu(n)\mu(n-k)$$
$$k = 0, \pm 1, \pm 2, \dots$$

当期望为 0 时，有 $c(n, n-k) = r(n, n-k)$

5. 广义平稳：均值与时间无关，即 $E(\mu(n)) = \mu$ ，其自相关函数、自协方差函数与时间无关，只与样值之间的时间差有关。即： $r(n, n-k) = r(k)$ ， $c(n, n-k) = c(k)$ 。

6. 平均各态历经（遍历性）：对于广义平稳的随机过程，其均值和相关函数具有各态历经性，也称具有遍历性。具有具有各态历经性的随机序列，可以用时间平均来获得集平均（期望）。

随机过程的 N 个样本的时间平均为：

$$\hat{\mu}_N = \frac{1}{N} \sum_{n=0}^{N-1} u(n)$$

对于所有的 N 值，可得 $E[\hat{\mu}_N] = \mu$ ，当 $N \rightarrow \infty$ 时，有 $\hat{\mu}_N = \mu$ ，则称随机过程在均值意义上是各态历经的，因为概率分布不随时间发生变化。

6. 相关矩阵：

定义随机向量 $\mathbf{u}(n) = [u(n), u(n-1), \dots, u(n-M+1)]^T$ ，则该随机过程的 M 为相关矩阵的定义为：

复数域： $\mathbf{R} = E[\mathbf{u}(n)\mathbf{u}^H(n)]$ （上标 H 表示转置及共轭操作）

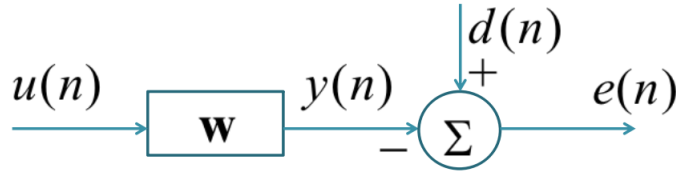
实数域： $\mathbf{R} = E[\mathbf{u}(n)\mathbf{u}^T(n)]$

$$\mathbf{R} = \begin{bmatrix} r(0) & r(1) & \dots & r(M-1) \\ r(-1) & r(0) & \dots & r(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(-M+1) & r(-M+2) & \dots & r(0) \end{bmatrix}$$

复数域 $\mathbf{R}^H = \mathbf{R}$ ， $r(-k) = r^*(k)$

实数域 $\mathbf{R}^T = \mathbf{R}$ ， $r(-k) = r(k)$

2 维纳滤波器



维纳数字滤波框图

$$e(n) = d(n) - y(n)$$

已知输入信号 $u(n)$ 和期望响应(参考信号) $d(n)$ ，使误差信号 $e(n)$ 在某种统计意义上最小。

2.1 正交性原理

对于因果滤波器，其 n 时刻的输出为卷积：

$$y(n) = \sum_{k=0}^{\infty} w_k^* u(n-k) = \mathbf{w}^H \mathbf{u}(n), \quad n = 0, 1, 2, \dots$$

使用均方误差作为代价函数，优点：该代价函数具有唯一的最小值。

$$J(n) = E[|e(n)|^2] = E[e(n)e^*(n)]$$

如何求最小值：求函数对所有变量的偏导数，令所有偏导数为 0，其对应的变量为多元函数取得最小值的解。为了方便表示，常将所有的偏导数写成列向量，该列向量称为函数的梯度或共轭梯度。因此，令所有偏导数为 0，等价于令梯度或共轭梯度为 0。

第 k 个滤波器的系数可表示为：

$$w_k = a_k + jb_k, \quad k = 0, 1, 2, \dots$$

整个滤波器系数向量可表示为：

$$\mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_{M-1} \end{bmatrix} = \begin{bmatrix} a_0 + jb_0 \\ a_1 + jb_1 \\ \vdots \\ a_{M-1} + jb_{M-1} \\ \vdots \end{bmatrix}$$

把代价函数 $J(n)$ 对 \mathbf{w} 中的每个 w_k 系数求偏导得到：

$$\nabla_k J = E \left[\frac{\partial e(n)}{\partial a_k} e^*(n) + \frac{\partial e^*(n)}{\partial a_k} e(n) + \frac{\partial e(n)}{\partial b_k} j e^*(n) + \frac{\partial e^*(n)}{\partial b_k} j e(n) \right], \quad k = 0, 1, 2, \dots$$

计算四个偏导数分别为：

$$\begin{aligned}\frac{\partial e(n)}{\partial a_k} &= -u(n-k), & \frac{\partial e^*(n)}{\partial a_k} &= -u^*(n-k) \\ \frac{\partial e(n)}{\partial b_k} &= ju(n-k), & \frac{\partial e^*(n)}{\partial b_k} &= -ju^*(n-k)\end{aligned}$$

将四个式子带入到 $\nabla_k J$ ，得到：

$$\nabla_k J = -2E[u(n-k)e^*(n)], \quad k = 0, 1, 2, \dots$$

当梯度为零时，代价函数达到最小值，所以梯度向量 ∇J 的所以元素都要等于 0：

$$\nabla_k J = -2E[u(n-k)e^*(n)] = 0, \quad k = 0, 1, 2, \dots$$

等效于 (下标 o 表示 optimal)：

$$E[u(n-k)e_o^*(n)] = 0$$

正交性原理：代价函数最小的充要条件是：估计误差 $e_o(n)$ 与 n 时刻进入期望响应估计的每个输入样值 $u(n-k)$ 正交。

正交性原理推论

根据自适应滤波器的输出，可得

$$\begin{aligned}E[y(n)e^*(n)] &= E\left[\sum_{k=0}^{\infty} w_k^* u(n-k)e^*(n)\right] \\ &= \sum_{k=0}^{\infty} w_k^* E[u(n-k)e^*(n)]\end{aligned}$$

根据正交性原理

$$E[u(n-k)e_o^*(n)] = 0, \quad k = 0, 1, 2, \dots$$

$$\text{可推得 } E[y_o(n)e_o^*(n)] = 0$$

2.2 最小均方误差

达到最优时：

$$e_o(n) = d(n) - y_o(n) = d(n) - \hat{d}(n|\mathcal{U}_n)$$

\mathcal{U}_n 表示输入限号 $u(n)$ 直到时刻 n 的输入样值张成的空间。根据卷积公式，计算得到的输出实际上就是直到时刻 n 的输入样值的线性组合。

$$d(n) = \hat{d}(n|\mathcal{U}_n) + e_o(n)$$

$$d^*(n) = \hat{d}^*(n|\mathcal{U}_n) + e_o^*(n)$$

所以：

$$\sigma_d^2 = d(n)d^*(n) = \sigma_{\hat{d}(n|\mathcal{U}_n)}^2 + E[\hat{d}(n|\mathcal{U}_n)e_o^*(n)] + E[\hat{d}^*(n|\mathcal{U}_n)e_o(n)] + \sigma_{e_o}^2$$

由正交性原理的推论可得，式子中的两项期望为 0，所以有：

$$\sigma_d^2 = \sigma_{\hat{d}}^2 + J_{min}$$

$$J_{min} = \sigma_d^2 - \sigma_{\hat{d}}^2$$

归一化均方误差：

$$\sigma = \frac{J_{min}}{\sigma_d^2} = 1 - \frac{\sigma_{\hat{d}}^2}{\sigma_d^2}, \quad 0 \leq \sigma \leq 1$$

2.3 维纳-霍夫方程（求解）

一般情况下的维纳-霍夫方程

根据最优情况下误差信号的定义

$$e_o(n) = d(n) - y_o(n) = d(n) - \sum_{i=0}^{\infty} w_{o,i}^* u(n-i)$$

和正交性原理

$$E[u(n-k)e_o^*(n)] = 0, \quad k = 0, 1, 2, \dots$$

可得

$$\sum_{i=0}^{\infty} w_{o,i} E[u(n-k)u^*(n-i)] = E[u(n-k)d^*(n)]$$

$$k = 0, 1, 2, \dots$$

第一个期望 $E[u(n-k)u^*(n-i)]$ ：等于相隔 $i-k$ 个延迟的输入信号自相关函数，即

$$r(i-k) = E[u(n-k)u^*(n-i)]$$

第二个期望 $E[u(n-k)d^*(n)]$ ：等于输入信号 $u(n-k)$ 与期望响应 $d(n)$ 相隔 $-k$ 个延迟的互相关，即

$$p(-k) = E[u(n-k)d^*(n)]$$

将 $r(i-k)$ 与 $p(-k)$ 代入：

可得维纳霍-夫方程

$$\sum_{i=0}^{\infty} w_{o,i} r(i-k) = p(-k), \quad k = 0, 1, 2, \dots$$

对于横向滤波器，第 n 时刻的输入向量可表示为

$$\mathbf{u}(n) = [u(n), u(n-1), \dots, u(n-M+1)]^T$$

则其相关矩阵为

$$\begin{aligned} \mathbf{R} &= E[\mathbf{u}(n)\mathbf{u}^H(n)] \\ &= \begin{bmatrix} r(0) & r(1) & \dots & r(M-1) \\ r^*(1) & r(0) & \dots & r(M-2) \\ \dots & \dots & \dots & \dots \\ r^*(M-1) & r^*(M-2) & \dots & r(0) \end{bmatrix} \end{aligned}$$

将滤波器的输入向量与期望响应的互相关向量记为

$$\mathbf{p} = E[\mathbf{u}(n)d^*(n)]$$

则维纳-霍夫方程可写成矩阵形式：

$$\mathbf{R}\mathbf{w}_o = \mathbf{p}$$

$$\text{其中， } \mathbf{w}_o = [w_{o,0}, w_{o,1}, \dots, w_{o,M-1}]^T$$

如果相关矩阵是非奇异的，则可解出令代价函数最小的最优权值向量，即

$$\mathbf{w}_o = \mathbf{R}^{-1}\mathbf{p}$$

换个角度：利用向量求导来求解

误差：

$$e(n) = d(n) - \mathbf{w}^H \mathbf{u}(n)$$

误差的平方：

$$\begin{aligned} e(n)e^*(n) &= (d(n) - \mathbf{w}^H \mathbf{u}(n))(d^*(n) - \mathbf{u}^H(n)\mathbf{w}) \\ &= |d(n)|^2 - d(n)\mathbf{u}^H(n)\mathbf{w} - \mathbf{w}^H \mathbf{u}(n)d^*(n) + \mathbf{w}^H \mathbf{u}(n)\mathbf{u}^H(n)\mathbf{w} \end{aligned}$$

对误差的平方取期望：

$$E[|e(n)|^2] = E[|d(n)|^2] + E[d(n)\mathbf{u}^H(n)]\mathbf{w} + \mathbf{w}^H E[\mathbf{u}(n)d^*(n)] + \mathbf{w}^H E[\mathbf{u}(n)\mathbf{u}^H(n)]\mathbf{w}$$

其中 $E[\mathbf{u}(n)\mathbf{u}^H(n)] = \mathbf{R}$ ，并令 $E[\mathbf{u}(n)d^*(n)] = \mathbf{p}$ （与维纳-霍夫方程中的记法相同），带入上式得损失函数：

$$J(\mathbf{w}) = E[|e(n)|^2] = \sigma_d^2 - \mathbf{p}^H \mathbf{w} + \mathbf{w}^H \mathbf{p} + \mathbf{w}^H \mathbf{R} \mathbf{w}$$

对上式求 \mathbf{w} 的共轭梯度，并令其等于 0：

$$\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}^*} = -\mathbf{p} + \mathbf{R}\mathbf{w} = 0$$

同样可以求得参数 \mathbf{w} 。

3 最速下降法

3.1 最速下降法的基本思想

迭代下降法:从某一初始值 $\mathbf{w}(0)$ 开始,按照固定的步骤,产生一系列权重向量 $\mathbf{w}(1), \mathbf{w}(2), \mathbf{w}(3), \dots$, 使得代价函数的值在每一次迭代之后都下降:

$$J(\mathbf{w}(n+1)) < J(\mathbf{w}(n))$$

迭代下降法的一种简单形式——最速下降法,沿着负梯度方向连续调整权重向量 \mathbf{w} , 将梯度表示为:

$$\mathbf{g} = \nabla J(\mathbf{w})$$

从而最速下降法可表示为 (μ 为步长):

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \frac{1}{2}\mu\mathbf{g}(n)$$

3.2 最速下降算法应用于维纳滤波器

第 2 章中提到维纳滤波器的代价函数为:

$$J(\mathbf{w}) = E[|e(n)|^2] = \sigma_d^2 - \mathbf{p}^H \mathbf{w}(n) + \mathbf{w}^H(n) \mathbf{p} + \mathbf{w}^H(n) \mathbf{R} \mathbf{w}(n)$$

其中:

$$\sigma_d^2 = E[d^2(n)]$$

$$\mathbf{p} = E[\mathbf{u}(n)d^*(n)]$$

$$\mathbf{R} = E[\mathbf{u}(n)\mathbf{u}^H(n)]$$

其梯度向量为:

$$\nabla J(n) = -2\mathbf{p} + 2\mathbf{R}\mathbf{w}(n)$$

代入最速下降法的公式中, 迭代解为:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu(\mathbf{p} - \mathbf{R}\mathbf{w}(n))$$

【1】 与维纳滤波器的闭合解 $\mathbf{w}_o = \mathbf{R}^{-1}\mathbf{p}$ 相比, 迭代解不要求相关矩阵 \mathbf{R} 的逆。

【2】 迭代解是经典的最小均方算法的基础。

3.3 最速下降法的稳定性

问: 当 $n \rightarrow \infty$ 时, 是否有 $\mathbf{w}(n) \rightarrow \mathbf{w}_o$? 若有, 需要满足什么条件?

定义: n 时刻的权重误差向量:

$$\mathbf{c}(n) = \mathbf{w}_o - \mathbf{w}(n)$$

将迭代式:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu(\mathbf{p} - \mathbf{R}\mathbf{w}(n))$$

两边同时减去 \mathbf{w}_o ，并消去负号，可得：

$$\mathbf{c}(n+1) = \mathbf{c}(n) - \mu(\mathbf{p} - \mathbf{R}\mathbf{w}(n))$$

将维纳方程 $\mathbf{R}\mathbf{w}_o = \mathbf{p}$ 代入上式，消去 \mathbf{p} ，可得：

$$\mathbf{c}(n+1) = \mathbf{c}(n) - \mu\mathbf{R}(\mathbf{w}_o - \mathbf{w}(n))$$

$$\mathbf{c}(n+1) = (\mathbf{I} - \mu\mathbf{R})\mathbf{c}(n)$$

根据上式的误差向量的关系类推，得到：

$$\mathbf{c}(n+1) = (\mathbf{I} - \mu\mathbf{R})^2\mathbf{c}(n-1)$$

$$\mathbf{c}(n+1) = (\mathbf{I} - \mu\mathbf{R})^{n+1}\mathbf{c}(0)$$

每一次迭代之后误差向量都乘上了 $(\mathbf{I} - \mu\mathbf{R})$ ，如果该项小于 0，那么误差向量是在不断减小的。

使用特征值分解，将相关矩阵分解为：

$$\mathbf{R} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H$$

迭代式变为：

$$\mathbf{c}(n+1) = (\mathbf{I} - \mu\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H)\mathbf{c}(n)$$

两边同乘以 \mathbf{Q}^H ：

$$\begin{aligned}\mathbf{Q}^H\mathbf{c}(n+1) &= (\mathbf{Q}^H - \mu\mathbf{\Lambda}\mathbf{Q}^H)\mathbf{c}(n) \\ &= (\mathbf{I} - \mu\mathbf{\Lambda})\mathbf{Q}^H\mathbf{c}(n)\end{aligned}$$

定义变换向量 $\mathbf{v}(n) = \mathbf{Q}^H\mathbf{c}(n)$ ，代换后有：

$$\mathbf{v}(n+1) = (\mathbf{I} - \mu\mathbf{\Lambda})\mathbf{v}(n)$$

初始权重常取 $\mathbf{w}(0) = \mathbf{0}$ ，所以：

$$\mathbf{v}(0) = \mathbf{Q}^H\mathbf{c}(0) = \mathbf{Q}^H[\mathbf{w}_o - \mathbf{0}] = \mathbf{Q}^H\mathbf{w}_o$$

➤ 若证明 $n \rightarrow \infty$ 时， $\mathbf{w}(n) \rightarrow \mathbf{w}_o$

可证明 $n \rightarrow \infty$ 时， $\mathbf{c}(n) = \mathbf{w}_o - \mathbf{w}(n) \rightarrow \mathbf{0}$

或证明 $n \rightarrow \infty$ 时， $\mathbf{v}(n) = \mathbf{Q}^H\mathbf{c}(n) \rightarrow \mathbf{0}$

➤ 令 $v_k(n)$ 为 $\mathbf{v}(n)$ 的第 k 个元素， $k = 1, 2, \dots, M$

即要求**证明**： $n \rightarrow \infty$ 时， $v_k(n+1) \rightarrow 0$

➤ 式 $\mathbf{v}(n+1) = (\mathbf{I} - \mu\mathbf{\Lambda})\mathbf{v}(n)$ 的展开式为

$$\begin{bmatrix} v_1(n+1) \\ v_2(n+1) \\ \vdots \\ v_M(n+1) \end{bmatrix} = \begin{bmatrix} 1-\mu\lambda_1 & & & \\ & 1-\mu\lambda_2 & & \\ & & \ddots & \\ & & & 1-\mu\lambda_M \end{bmatrix} \begin{bmatrix} v_1(n) \\ v_2(n) \\ \vdots \\ v_M(n) \end{bmatrix}$$

$$v_1(n+1) = (1 - \mu\lambda_1)v_1(n)$$

$$v_2(n+1) = (1 - \mu\lambda_2)v_2(n)$$

➤ 或写成

$$\vdots$$

$$v_M(n+1) = (1 - \mu\lambda_M)v_M(n)$$

➤ 由第 k 个方程

$$v_k(n+1) = (1 - \mu\lambda_k)v_k(n), \quad k = 1, 2, \dots, M$$

递推可得

$$v_k(n+1) = (1 - \mu\lambda_k)^{n+1} v_k(0)$$

➤ 要使得 $v_k(\infty) \rightarrow 0$ ，上式中必须满足

$$-1 < (1 - \mu\lambda_k) < 1, \quad k = 1, 2, \dots, M$$

➤ 因此，最速下降法的**稳定性条件**为 $0 < \mu < \frac{2}{\lambda_{\max}}$

其中

$$\lambda_{\max} = \max\{\lambda_1, \lambda_2, \dots, \lambda_M\}$$

4 自适应滤波器

上两章中求解维纳滤波器的封闭解和数值解时，需要的已知条件为：相关矩阵 \mathbf{R} ，输入信号与期望响应的互相关向量 \mathbf{p} 。

然而，在实际应用中， \mathbf{R} 和 \mathbf{p} 大部分情况下是未知的；如果信号是时变的，则 \mathbf{R} 和 \mathbf{p} 都是时变的。所以需要一种能够实时估计或者跟踪信号的统计特性方法，即自适应滤波方法。

4.1 LMS 自适应滤波算法

➤ 前面推导的最速下降算法迭代式为

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu[\mathbf{p} - \mathbf{R}\mathbf{w}(n)]$$

其中, $\mathbf{R} = E[\mathbf{u}(n)\mathbf{u}^H(n)]$, $\mathbf{p} = E[\mathbf{u}(n)d(n)]$

➤ 为了能够实时处理, 可以采用信号的瞬时值来近似估计, 即

$$\hat{\mathbf{R}}(n) = \mathbf{u}(n)\mathbf{u}^H(n), \hat{\mathbf{p}}(n) = \mathbf{u}(n)d(n)$$

➤ 将上式代入最速下降算法迭代式, 可得

$$\begin{aligned}\mathbf{w}(n+1) &= \mathbf{w}(n) + \mu \left[\overbrace{\mathbf{u}(n)d^*(n)}^{\hat{\mathbf{p}}(n)} - \overbrace{\mathbf{u}(n)\mathbf{u}^H(n)\mathbf{w}(n)}^{\hat{\mathbf{R}}(n)} \right] \\ &= \mathbf{w}(n) + \mu \mathbf{u}(n) \left[d^*(n) - \mathbf{u}^H(n)\mathbf{w}(n) \right] \\ &= \mathbf{w}(n) + \mu \mathbf{u}(n) \left[d(n) - \overbrace{\mathbf{w}^H(n)\mathbf{u}(n)}^{y(n)} \right]^* \\ &= \mathbf{w}(n) + \mu \mathbf{u}(n)e^*(n)\end{aligned}$$

上式即为**最小均方(LMS)算法**的迭代公式。

最小均方(LMS)算法的迭代公式, 包含三个过程:

- 1) 滤波输出: $y(n) = \mathbf{w}^H(n)\mathbf{u}(n)$
- 2) 计算估计误差: $e(n) = d(n) - y(n)$
- 3) 权值向量自适应: $\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \mathbf{u}(n)e^*(n)$

4.2 LMS 算法的收敛性证明

证明参数是否收敛到某一个值:

➤ 将LMS算法的迭代公式展开，可得

$$\begin{aligned}\mathbf{w}(n+1) &= \mathbf{w}(n) + \mu[\mathbf{u}(n)e^*(n)] \\ &= \mathbf{w}(n) + \mu\{\mathbf{u}(n)[d^*(n) - \mathbf{u}^H(n)\mathbf{w}(n)]\}\end{aligned}$$

➤ 上式两边取期望，可得

$$\begin{aligned}E[\mathbf{w}(n+1)] &= E[\mathbf{w}(n)] + \mu E[\mathbf{u}(n)d^*(n)] \\ &\quad - \mu E[\mathbf{u}(n)\mathbf{u}^H(n)\mathbf{w}(n)]\end{aligned}$$

近似处理： $\mathbf{w}(n)$ 与 $\mathbf{u}(n)$ 不相关：

➤ 在步长取值很小的情况下， $\mathbf{w}(n)$ 变化很慢，可以近似为与输入向量 $\mathbf{u}(n)$ 不相关，则近似有

$$E[\mathbf{u}(n)\mathbf{u}^H(n)\mathbf{w}(n)] = E[\mathbf{u}(n)\mathbf{u}^H(n)]E[\mathbf{w}(n)]$$

➤ 于是，迭代式

$$\begin{aligned}E[\mathbf{w}(n+1)] &= E[\mathbf{w}(n)] + \mu E[\mathbf{u}(n)d^*(n)] \\ &\quad - \mu E[\mathbf{u}(n)\mathbf{u}^H(n)\mathbf{w}(n)]\end{aligned}$$

可简化成

$$E[\mathbf{w}(n+1)] = (\mathbf{I} - \mu\mathbf{R})E[\mathbf{w}(n)] + \mu\mathbf{p}$$

➤ 设初始权值为 $\mathbf{w}(0)$ ，则有上式递推可得

$$\begin{aligned}E[\mathbf{w}(1)] &= (\mathbf{I} - \mu\mathbf{R})E[\mathbf{w}(0)] + \mu\mathbf{p} \\ E[\mathbf{w}(2)] &= (\mathbf{I} - \mu\mathbf{R})E[\mathbf{w}(1)] + \mu\mathbf{p} \\ &= (\mathbf{I} - \mu\mathbf{R})^2 E[\mathbf{w}(0)] + \mu(\mathbf{I} - \mu\mathbf{R})\mathbf{p} + \mu\mathbf{p} \\ &\vdots\end{aligned}$$

$$E[\mathbf{w}(n+1)] = (\mathbf{I} - \mu\mathbf{R})^{n+1} E[\mathbf{w}(0)] + \mu \sum_{i=0}^n (\mathbf{I} - \mu\mathbf{R})^i \mathbf{p}$$

将 \mathbf{R} 对角化：

➤ 根据特征值分解，可令 $\mathbf{R} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H$

其中， \mathbf{Q} 为酉矩阵，满足 $\mathbf{Q}\mathbf{Q}^H = \mathbf{I}$, $\mathbf{Q}^H = \mathbf{Q}^{-1}$

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \vdots \\ 0 & & & \lambda_M \end{pmatrix}$$

➤ 则递推式

$$E[\mathbf{w}(n+1)] = (\mathbf{I} - \mu\mathbf{R})^{n+1} E[\mathbf{w}(0)] + \mu \sum_{i=0}^n (\mathbf{I} - \mu\mathbf{R})^i \mathbf{p}$$

可转化为

$$\begin{aligned} E[\mathbf{w}(n+1)] &= (\mathbf{I} - \mu\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1})^{n+1} E[\mathbf{w}(0)] \\ &\quad + \mu \sum_{i=0}^n (\mathbf{I} - \mu\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1})^i \mathbf{p} \end{aligned}$$

单位矩阵拆分， $\mathbf{Q}, \mathbf{Q}^{-1}$ 左右抵消：

➤ 根据酉矩阵的性质 $\mathbf{Q}\mathbf{Q}^H = \mathbf{I}$, $\mathbf{Q}^H = \mathbf{Q}^{-1}$ ，递推式

$$\begin{aligned} E[\mathbf{w}(n+1)] &= (\mathbf{I} - \mu\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1})^{n+1} E[\mathbf{w}(0)] \\ &\quad + \mu \sum_{i=0}^n (\mathbf{I} - \mu\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1})^i \mathbf{p} \end{aligned}$$

等号右边第一项的矩阵可写为

$$\begin{aligned} (\mathbf{I} - \mu\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1})^n &= (\mathbf{Q}\mathbf{Q}^{-1} - \mu\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1})^n \\ &= [\mathbf{Q}(\mathbf{I} - \mu\mathbf{\Lambda})\mathbf{Q}^{-1}]^n \\ &= \mathbf{Q}(\mathbf{I} - \mu\mathbf{\Lambda})^n \mathbf{Q}^{-1} \end{aligned}$$

参数收敛，对步长的限制与最速下降法相同：

➤ 则递推式可转化为

$$E[\mathbf{w}(n+1)] = \mathbf{Q}(\mathbf{I} - \mu\mathbf{\Lambda})^{n+1}\mathbf{Q}^{-1}E[\mathbf{w}(0)] \\ + \mu\mathbf{Q}\sum_{i=0}^n(\mathbf{I} - \mu\mathbf{\Lambda})^i\mathbf{Q}^{-1}\mathbf{p}$$

要使上式收敛，必须满足如下条件：

$$|1 - \mu\lambda_{\max}| < 1 \Rightarrow 0 < \mu < \frac{2}{\lambda_{\max}}$$

然而特征值未知，近似处理：

➤ 条件 $0 < \mu < 2/\lambda_{\max}$ 需要知道相关矩阵的特征值，而实际应用中可能是未知的。

➤ 由于 $\lambda_{\max} < \sum_{i=0}^{M-1} \lambda_{i+1} = \text{tr}[\mathbf{R}] = \sum_{i=0}^{M-1} E[u^2(i)]$ ，则步长的范围可缩小为

$$0 < \mu < \frac{2}{\sum_{i=1}^N E[u_i^2]}$$

其中， $\sum_{i=0}^{M-1} E[u_i^2]$ 为输入信号的功率，通常可以估计。

参数收敛了，是否收敛到 \mathbf{w}_o ：

◆ 下面证明，当 $n \rightarrow \infty$ 时， $\mathbf{w}(n)$ 收敛于何处。

➤ 当 $n \rightarrow \infty$ 时，递推式

$$E[\mathbf{w}(n+1)] = \mathbf{Q}(\mathbf{I} - \mu\mathbf{\Lambda})^{n+1}\mathbf{Q}^{-1}E[\mathbf{w}(0)] \\ + \mu\mathbf{Q}\sum_{i=0}^n(\mathbf{I} - \mu\mathbf{\Lambda})^i\mathbf{Q}^{-1}\mathbf{p}$$

转化为

$$\lim_{n \rightarrow \infty} E[\mathbf{w}(n+1)] = \mu\mathbf{Q}\lim_{n \rightarrow \infty} \sum_{i=0}^n (\mathbf{I} - \mu\mathbf{\Lambda})^i \mathbf{Q}^{-1}\mathbf{p}$$

矩阵的等比数列求和（除法变为求逆）：

➤ 递推式中的矩阵 $\lim_{n \rightarrow \infty} \sum_{i=0}^n (\mathbf{I} - \mu \mathbf{\Lambda})^i$ 可转化为

$$\lim_{n \rightarrow \infty} \sum_{i=0}^n (\mathbf{I} - \mu \mathbf{\Lambda})^i = [\mathbf{I} - (\mathbf{I} - \mu \mathbf{\Lambda})]^{-1} = \frac{1}{\mu} \mathbf{\Lambda}^{-1}$$

将上式代入递推式

$$\lim_{n \rightarrow \infty} E[\mathbf{w}(n+1)] = \mu \mathbf{Q} \lim_{n \rightarrow \infty} \sum_{i=0}^n (\mathbf{I} - \mu \mathbf{\Lambda})^i \mathbf{Q}^{-1} \mathbf{p}$$

$$\begin{aligned} \text{可得 } \lim_{n \rightarrow \infty} E[\mathbf{w}(n+1)] &= \mu \mathbf{Q} \frac{1}{\mu} \mathbf{\Lambda}^{-1} \mathbf{Q}^{-1} \mathbf{p} = \mathbf{Q} \mathbf{\Lambda}^{-1} \mathbf{Q}^{-1} \mathbf{p} \\ &= \mathbf{Q} \mathbf{\Lambda}^{-1} \mathbf{Q}^H \mathbf{p} = \mathbf{R}^{-1} \mathbf{p} = \mathbf{w}_o \end{aligned}$$